

美团搜索中的用户引导

刘铭

2017-1-14

Outline



背景介绍

基础技术方案

多维度效果提升

总结与思考

背景介绍

搜索用户引导的What、Why、How

什么是用户引导？

引导

- 带着人们向某个目标行动，在行动上帮助人们走出困境。

搜索引导

- 帮助用户更快的完成搜索过程，找到目标信息。

具体到美团

- 帮助用户达成交易。

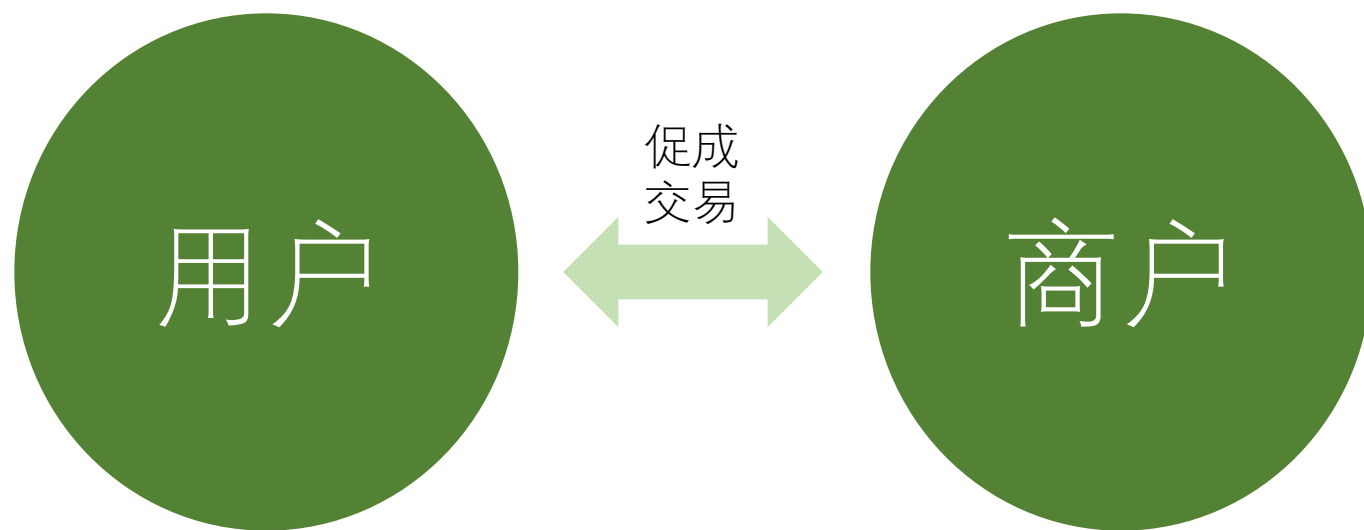


用户引导的作用

帮助用户获取所需，
帮助商户销售所售。

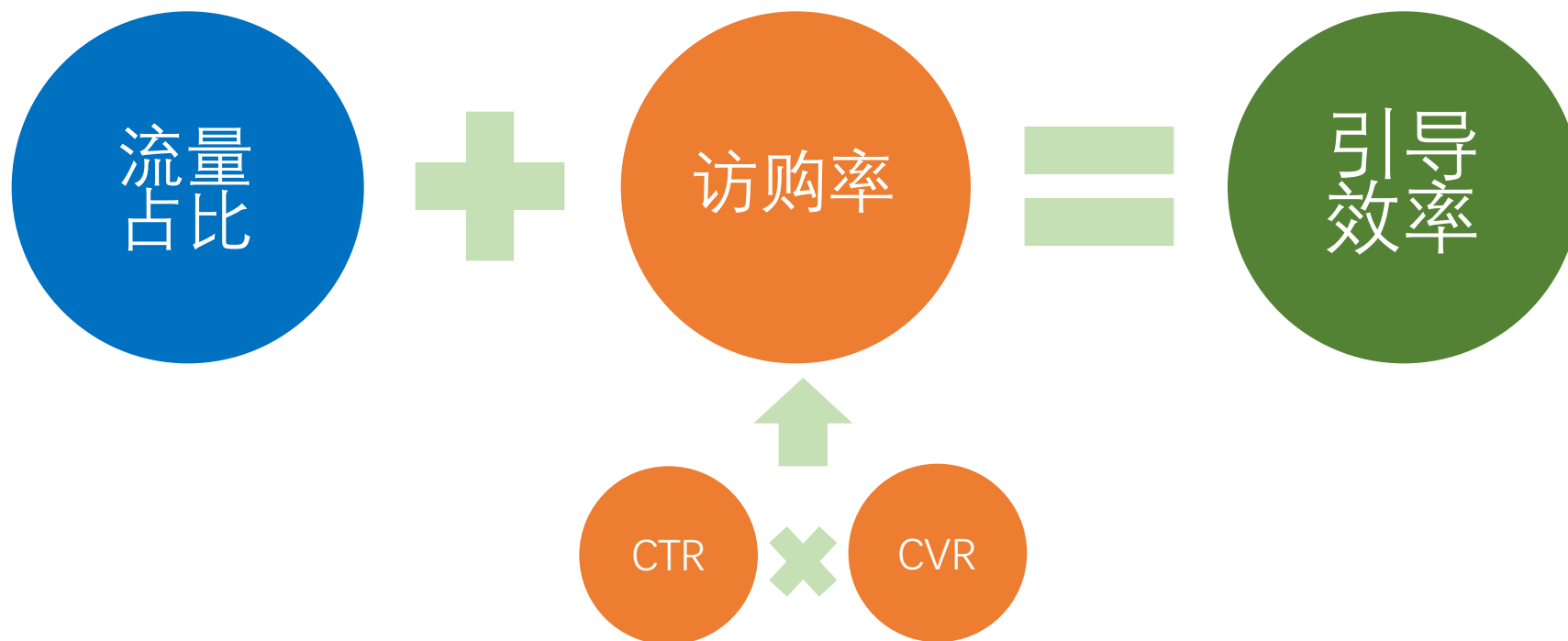
核心是用户。

- 辅助输入，提升易用性
- 辅助决策，提供信息
- 提升搜索体验
- 提升搜索质量



衡量指标

- 原则：不能是单边指标，需要存在制衡。
 - 非手工输入流量占比：体现了产品的易用性、智能化。
 - 搜索访购率：体现了引导的效率，与交易属性契合。



产品示例-搜索前



Click



热门搜索词

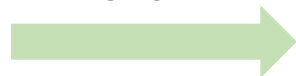
历史搜索词

产品示例-搜索中、搜索后

查询智能补全



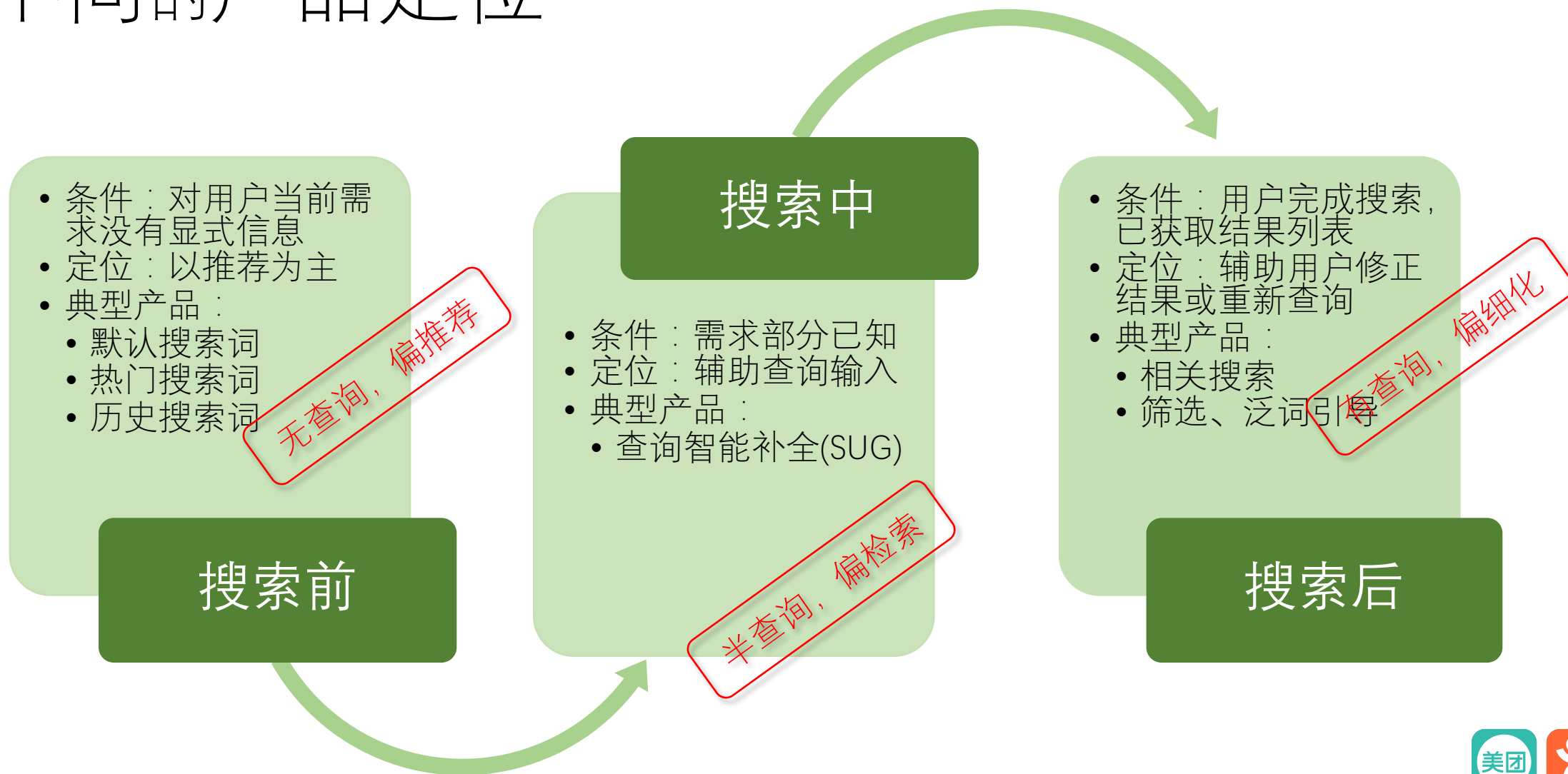
Click



泛词递进引导

相关搜索

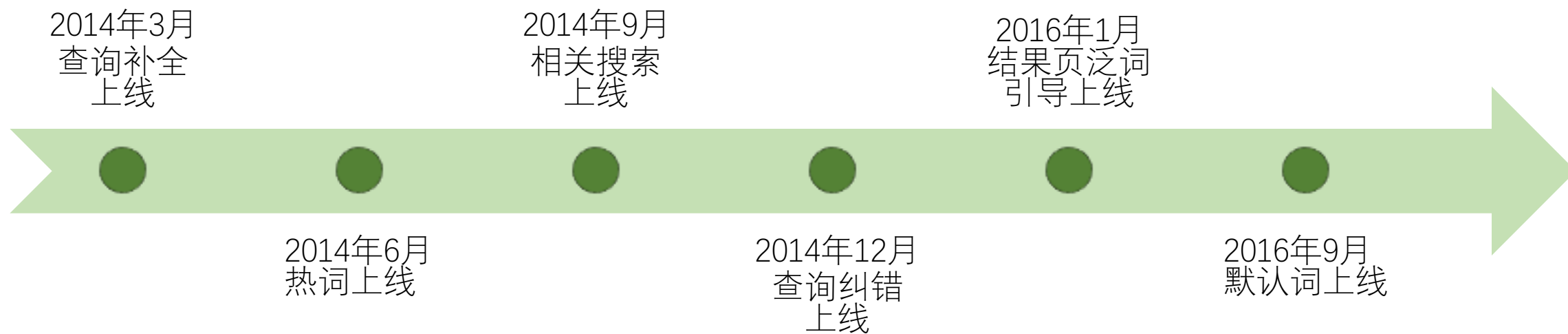
不同的产品定位



基础技术方案

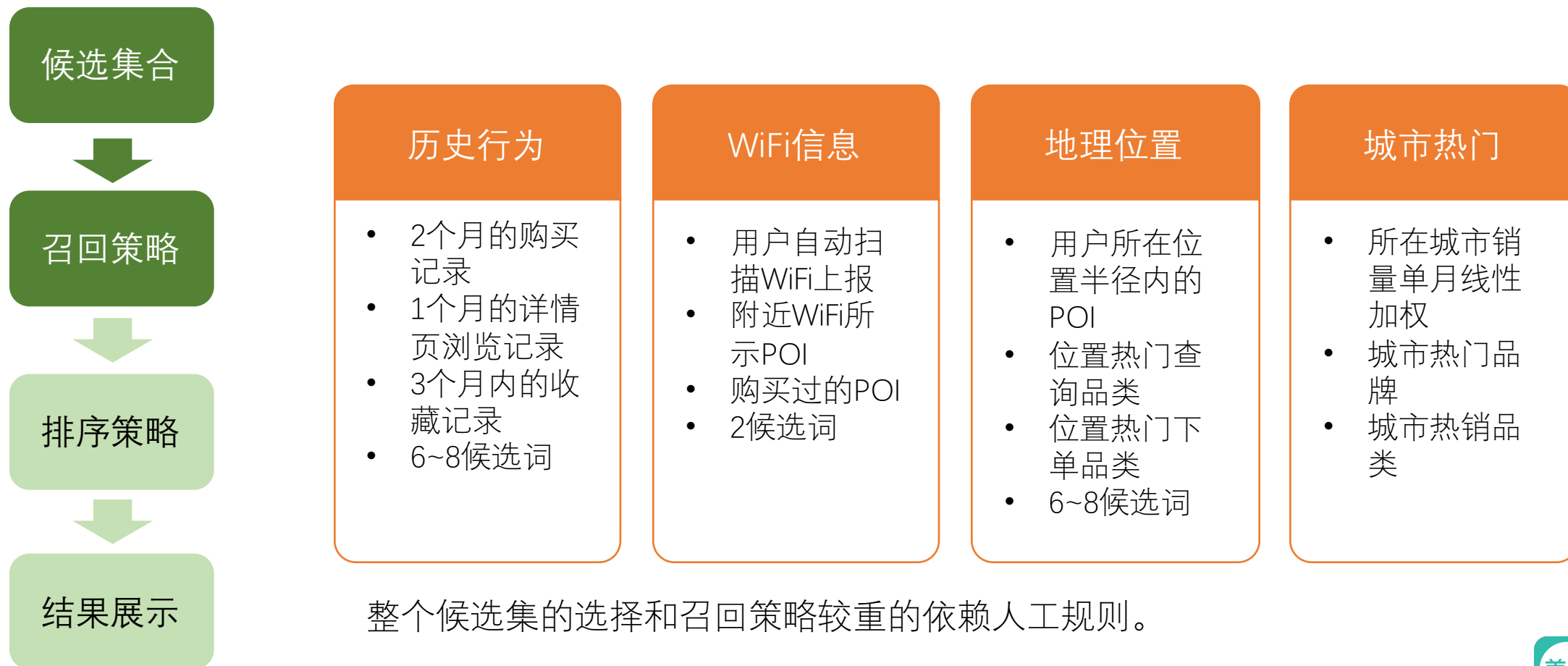
快速实现还可以用的版本

引导产品时间线



以热门搜索词和查询智能补全为例，看看构建的过程。

构建热词-1



整个候选集的选择和召回策略较重的依赖人工规则。

构建热词-2



展示采用了规则的九宫格，内容仅为文本。
默认搜索词作为热词的延伸，使用了类似的实现。

构建查询补全-1



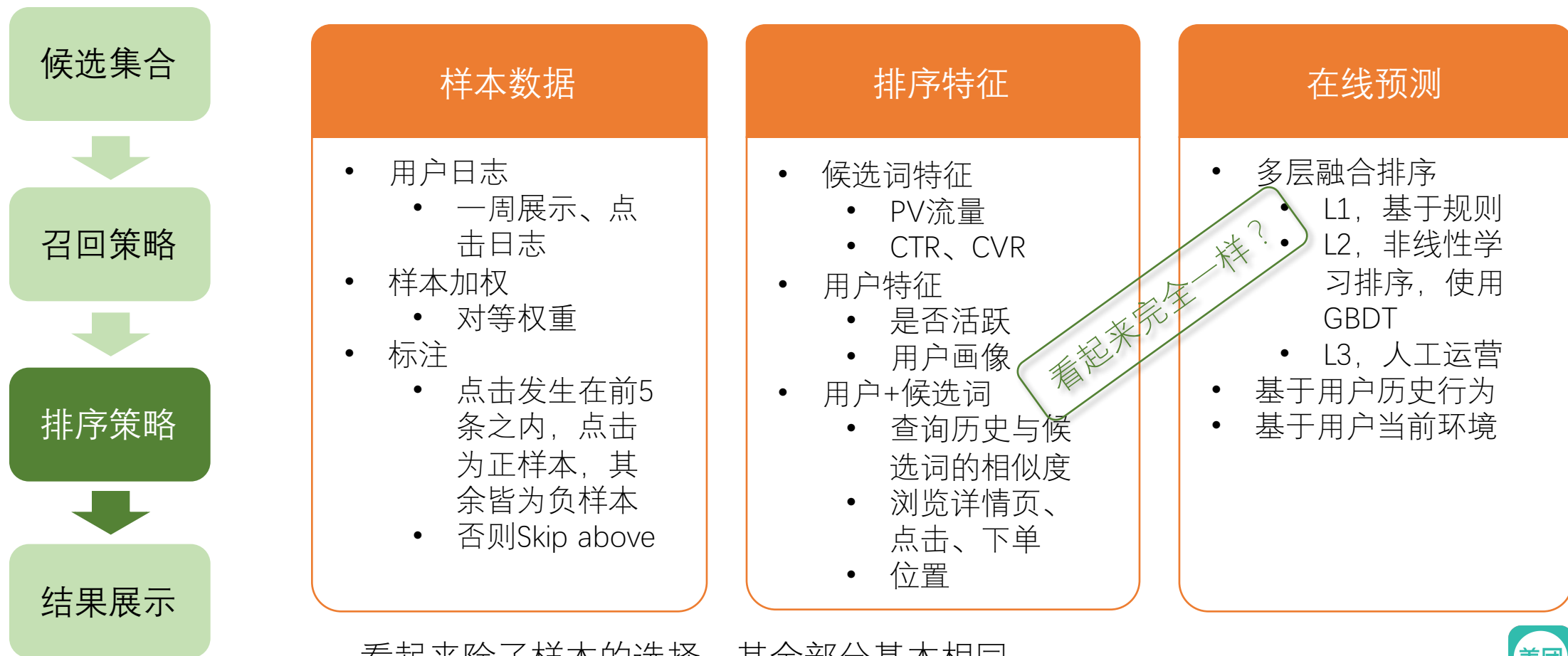
构建查询补全-2

使用前缀索引召回，是一种自然符合用户输入的方式。



优先级决定了排序的权重。分城市建立索引，先查本地，无结果扩展至全国。

构建查询补全-3



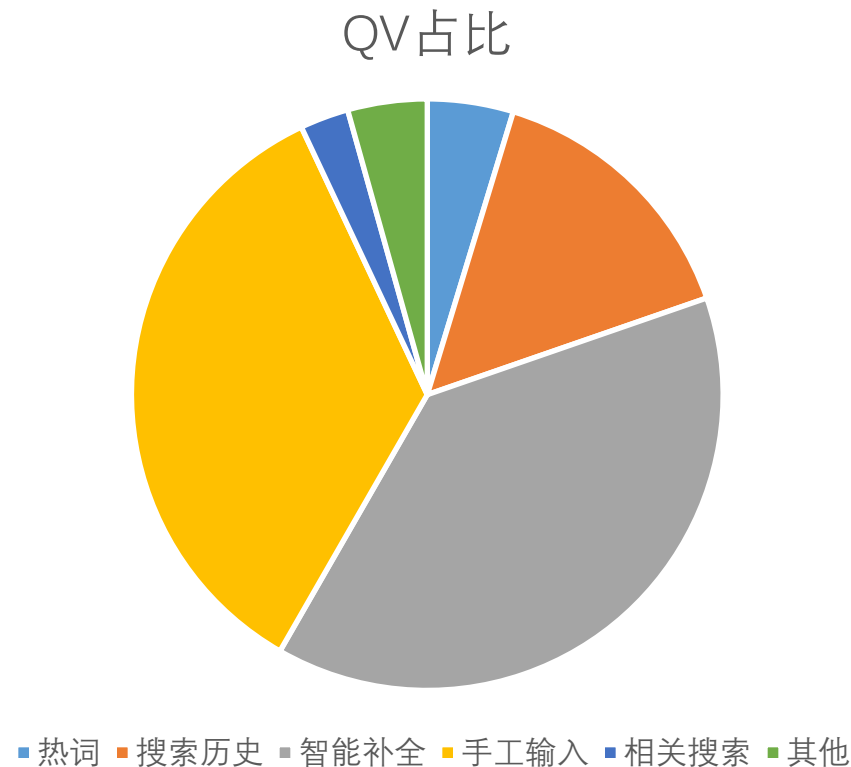
看起来除了样本的选择，其余部分基本相同。

存在问题

看起来功能都实现了，但是…

- 作为平台入口，需要服务多种业务，如何选择？CTR、CVR指向性太强。
- 多个功能相似又各有特点模块，多而杂乱的数据流程存，相似的个性化排序需求，存在冗余
- 多个独立服务，开发、运维成本高
- 是否真的满足了用户的需求？
- …

独立维护，流程冗余，效果一般。



多维度效果提升

提升迭代效率，尝试深度理解用户，将产品推向智能易用

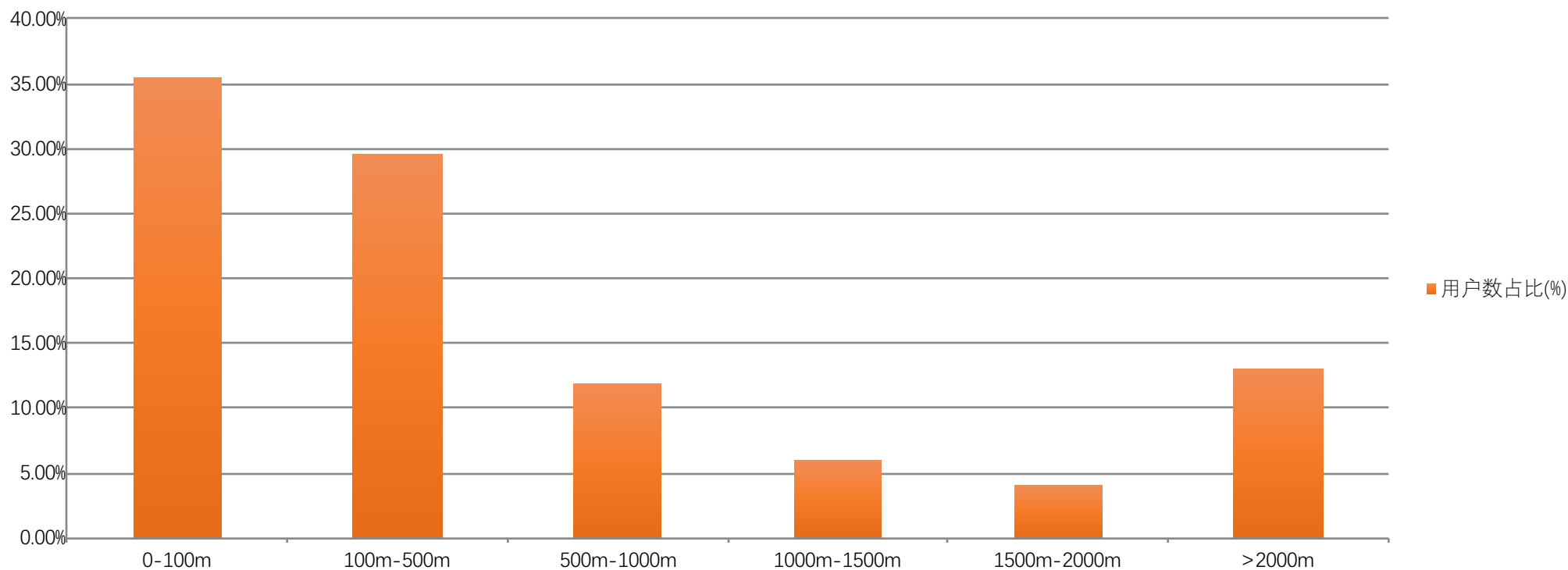
思考核心问题，我们做的是用户引导，核心是理解用户。

- 用户行为日志是金矿，做好数据分析，充分理解用户行为
- 在同行中有很多高手，看看大家是怎么面对同样的问题
- 换一个角度，站在用户的立场想一想
- 充分的数据分析后，有理有据，大胆假设

下单-消费 距离

87.02%用户下单与消费的距离<2000m，本地效应明显。

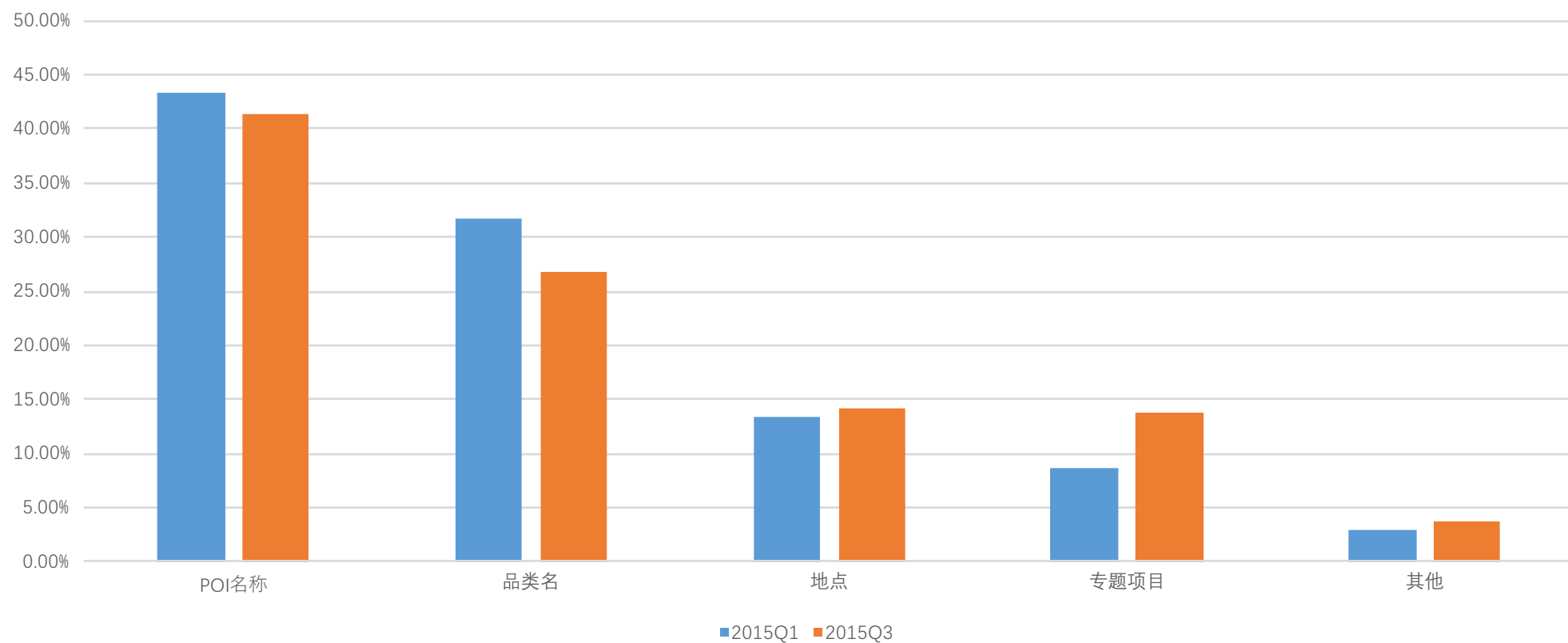
用户搜索-消费距离分布



查询类别分布

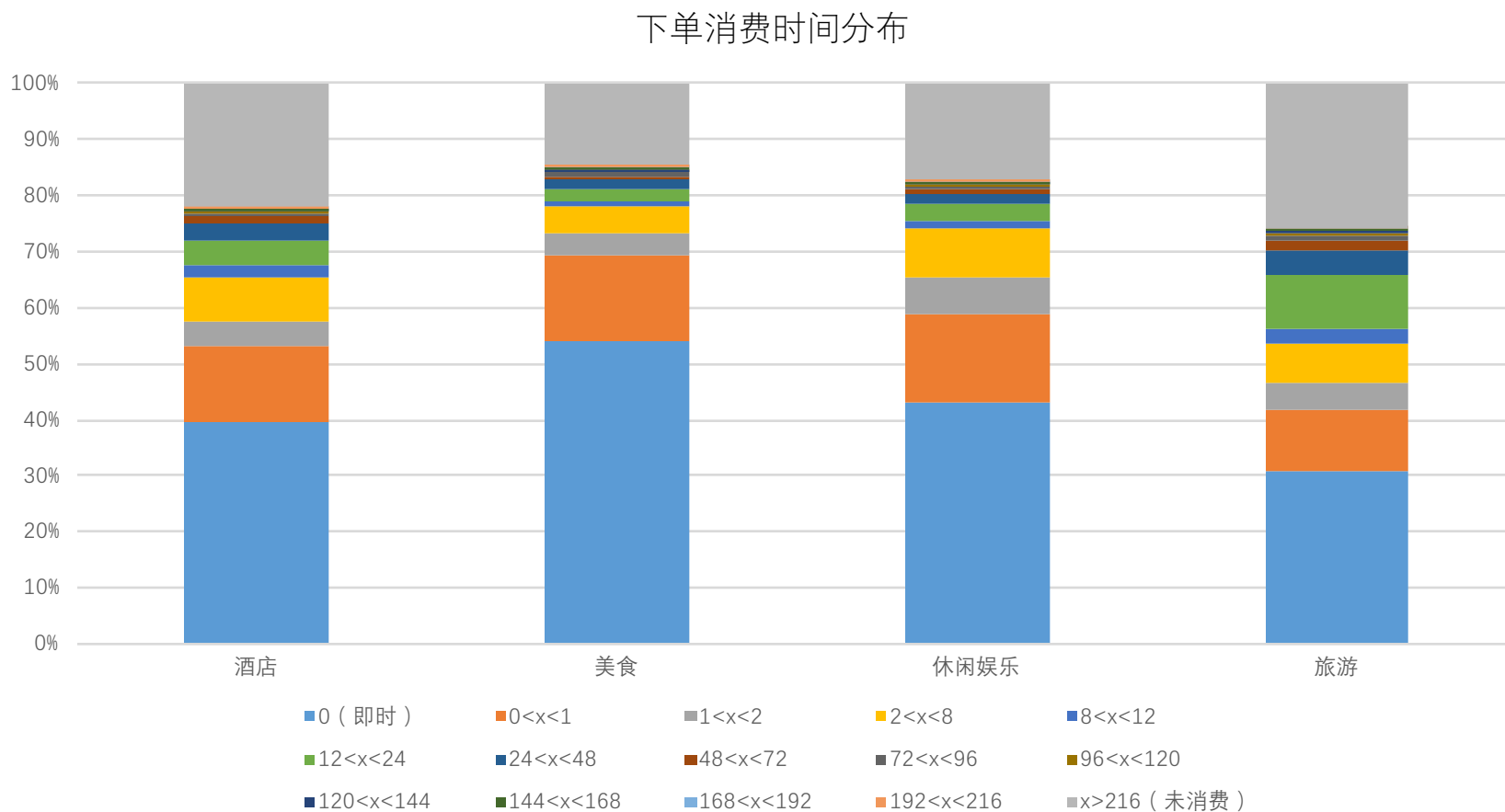
超过40%用户想找的就是明确需求的POI，交易属性非常强。

查询类别分布



下单-消费 时间

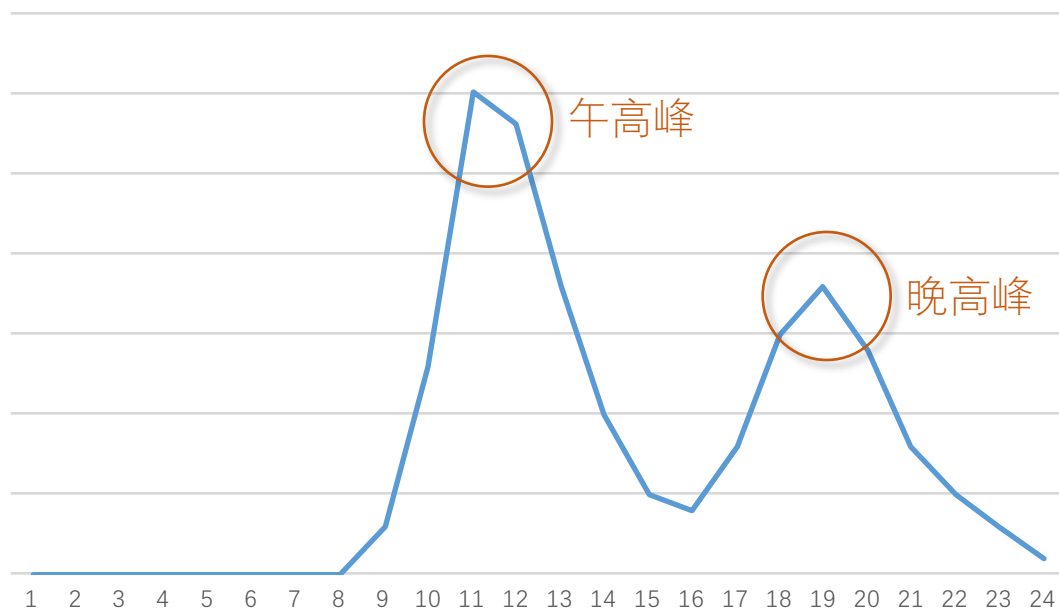
超过50%美食下单用户是即时消费，隐含着场景化信息，要求系统实时响应。



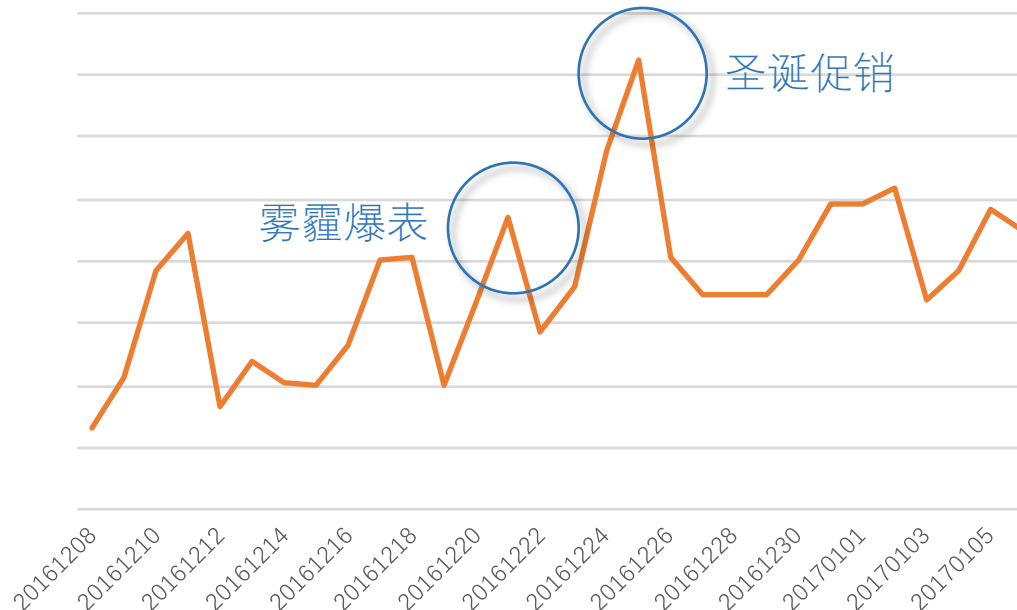
场景化特征分布

时间、天气、地点等Context特征影响决策，要求系统Context-aware。

美食外卖24时订单分布



外卖月订单趋势



O2O搜索的特点

本地化

大多数消费发生的距离半径不大，特别是外卖业务，配送距离就决定了候选集合。

个性化

千人千面，我搜索“烤鱼”，他也搜索“烤鱼”，我们的需求不一定相同。

美团搜索

时效性

大多数用户的消费都是即时行为，越来越多的在店即时消费，外卖业务的兴起。

场景化

环境特征对用户消费行为影响明显，如：时间、天气、场所、促销等。

大胆假设可能的改进

- POI直达
 - 既然用户POI搜索占比高，就给用户直达商家的路径？
 - 跳过搜索，会不会降低潜在的消费可能性？
- 瞬时意图预测
 - 默认搜索词和热词是无查询瞬时推荐场景，当前方案缺乏对实时意图的响应。
 - 构建另一个模型来预测用户打开APP的瞬时意图，结合场景预测意图？
- 泛查询递进引导
 - 品类、地点等意图不明确的查询类型，转化率很低，能否帮助用户明确所需？
 - 递进关系不正确，是否影响用户体验？
- 其他可能性…

仅是思想传达，
并非金科玉律。

实现POI直达

尝试**缩短决策路径**，跳过检索，直达目标。

- 强化地理位置特征
- 使用GeoHash召回预处理过的POI名称
- 与其他数据源召回结果归并排序



如何处理多源归并的排序问题？
轮替、瀑布流、还是交给模型？



POI直达在查询补全上应用-1

纯文本的查询补全



增加POI直达及业务图标

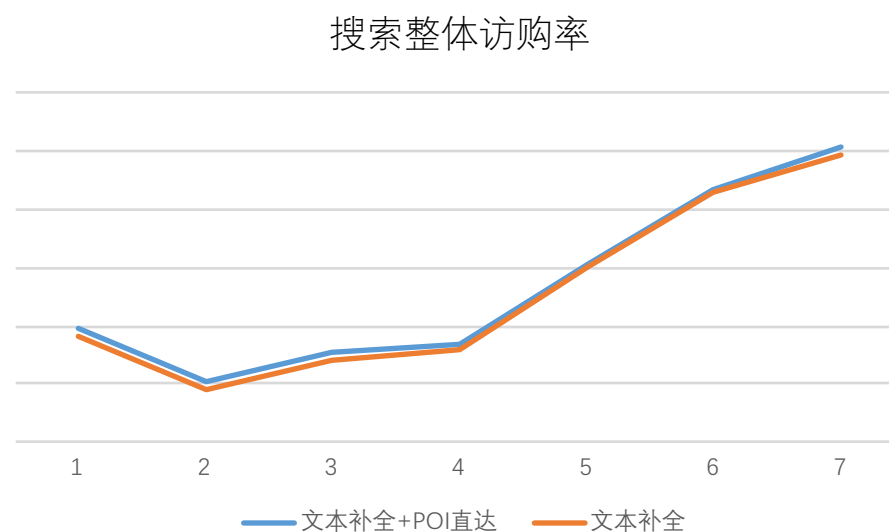
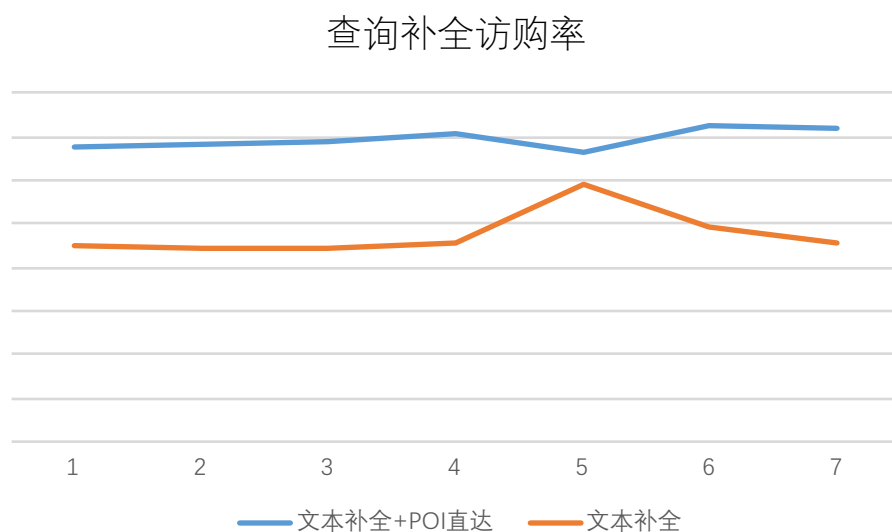


增加距离信息

增加业务图标

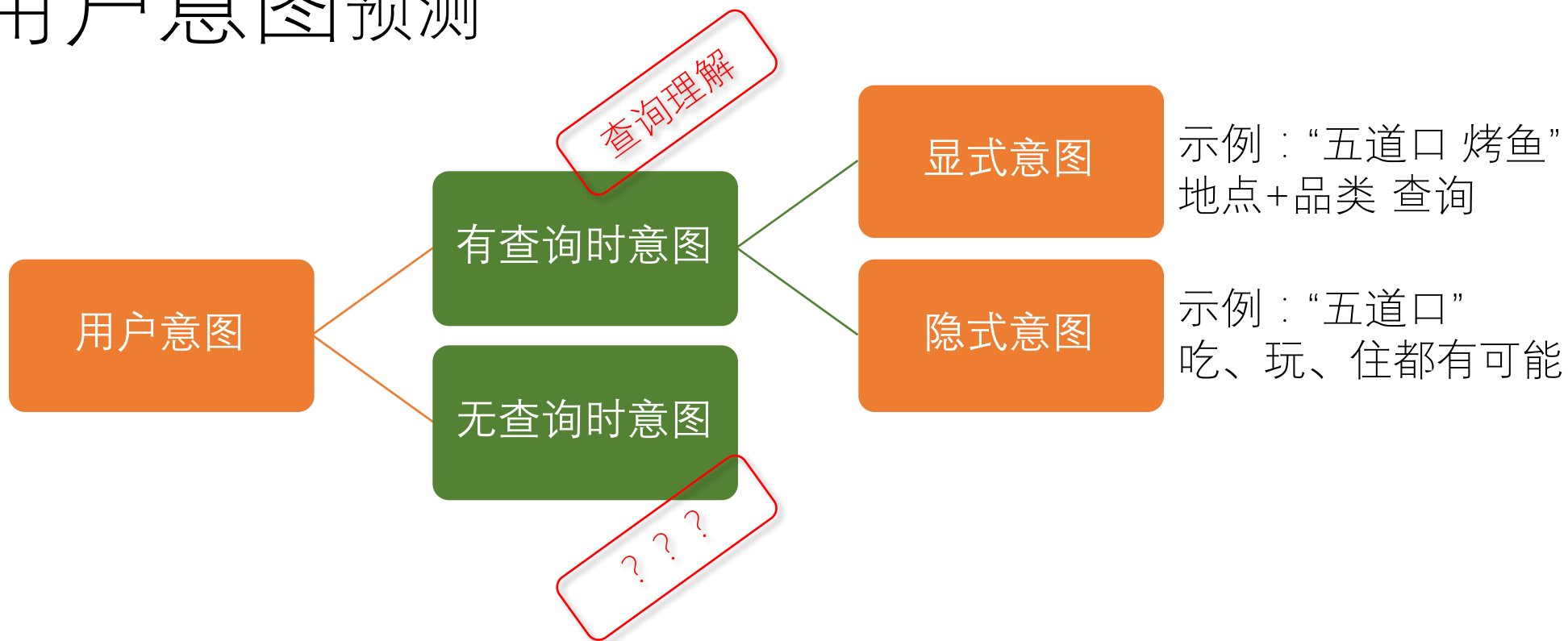
灵活使用图标、色彩表达产品的设计，增加决策信息，告诉用户可以这么玩。

POI直达在查询补全上应用-2



SUG单模块访购率和搜索整体访购率均有提升。

用户意图预测



默认搜索词和热门搜索词正是这样的场景，以往我们将用户历史特征直接加入学习排序模型，能否换一个思路，构建一个模型，**预测用户打开美团的瞬时**，想要找什么？

无查询瞬时意图预测

- 目标：建模预测用户打开APP时的瞬时意图
- 特点：实时性、无查询，本质上是多类别分布的求解



有多种方案可以选择：

- Softmax Regression
- Global Linear Model
- ...

考虑两类特征及其各种组合：

- Context：时间、天气、地点…
- User：查询、点击、下单…

关键是做好适合备选模型的特征处理。

我们的做法是将连续值特征、分类特征、数值特征都归一化到[0, 1]区间，注意查询、点击、下单特征要同时考虑时效性和累积性： $\text{score} = \alpha * \text{时效性} + (1 - \alpha) * \text{累积性}$ 。样本取自用户打开美团后第一次点击POI或是频道区的日志。

瞬时意图应用至默认搜索词

默认搜索词访购率季度趋势



对无查询瞬时意图单独建模预测本质上是做了一次模型Stacking

泛查询递进引导

泛查询的CVR很低，然而PV却很高…

典型的两类查询是**品类**和**地点**。前者是含义太泛，后者是意图太泛。

能否通过产品的属性细化品类查询，通过业务细化地点类查询？

定义为Tag，开放集合

定义为Tab，阶段闭集

注意控制展现，**有别于筛选**，重查询的结果需要保证。

Tags引导



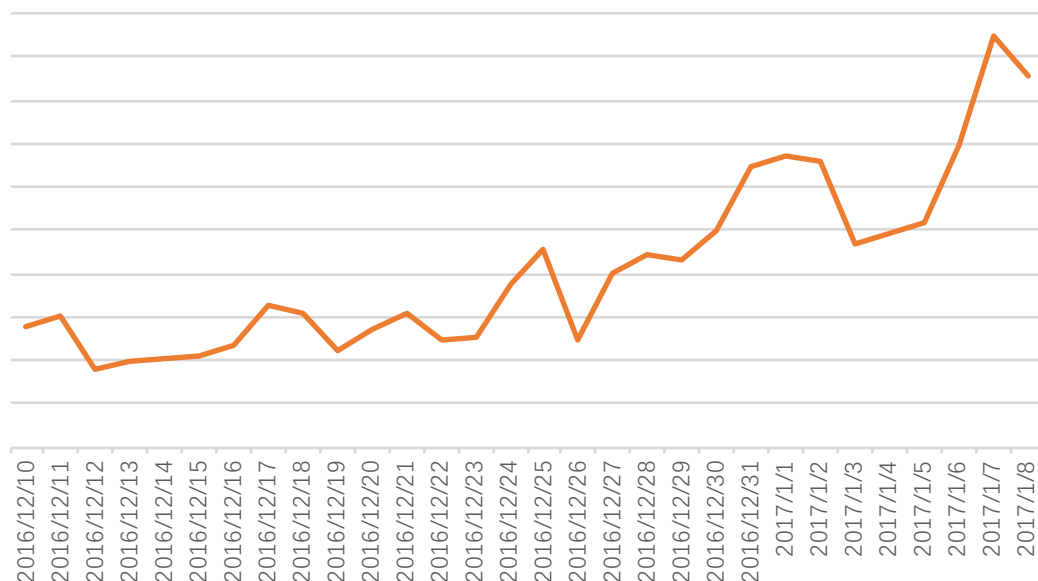
Tags引导



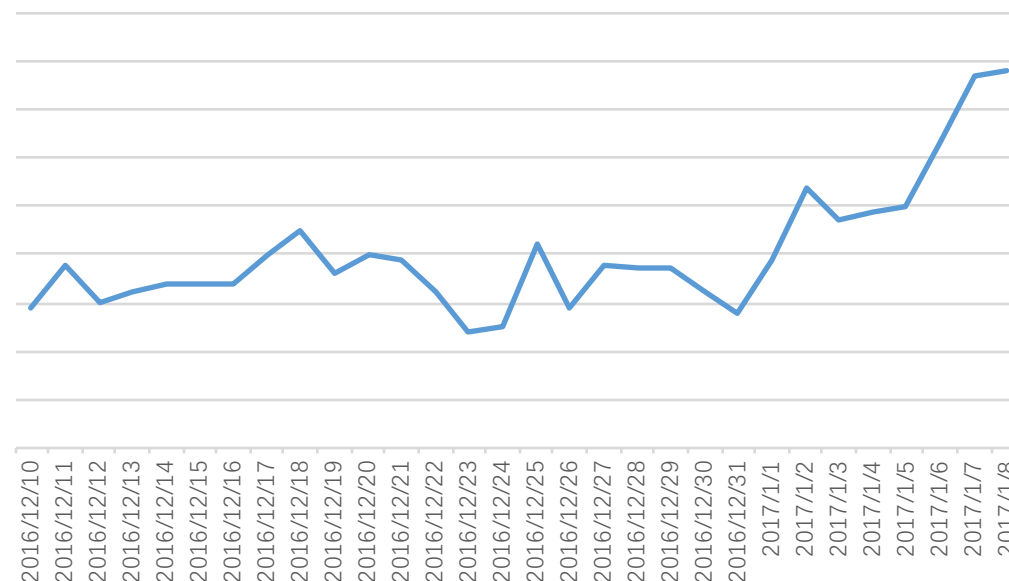
专题引导

泛查询递进引导的效果

月订单量趋势



月访购率趋势



新产品上线初期，用户积极尝试，热潮过了之后，指标开始回撤，持续优化，指标再次攀升。结果页的泛查询递进引导也使用了瞬时意图预测模型。

优化点组合应用



默认搜索词使用瞬时意图展示POI直达



热门搜索词使用瞬时意图展示POI直达

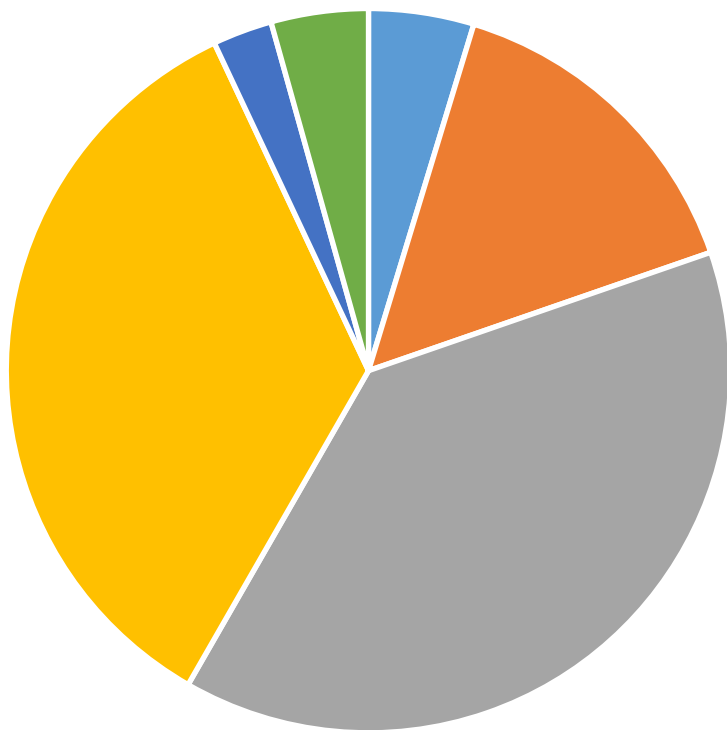
产品策略型优化和算法策略型优化可以尝试组合，触发新的场景。

持续优化细节

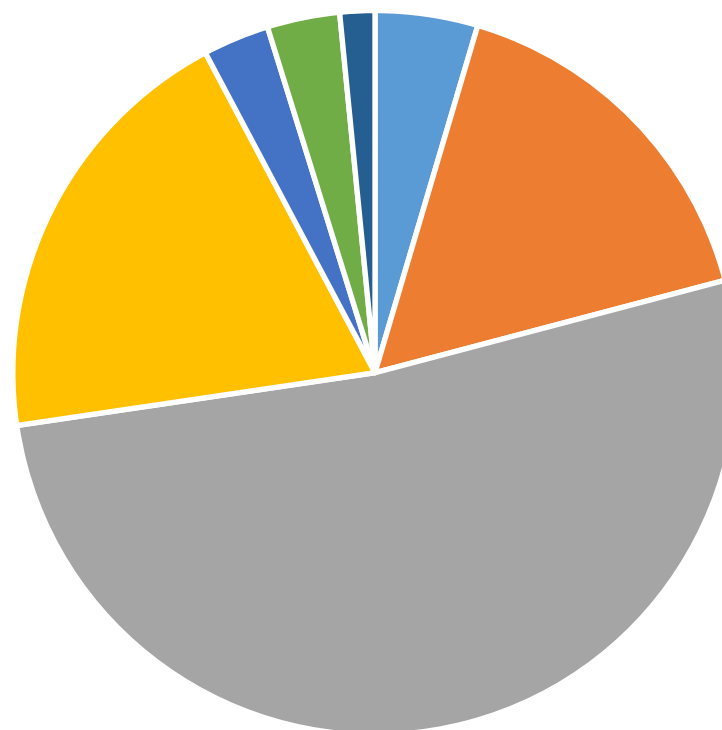
- 用户引导产品的优化是个细活，需要长期持久的投入。
 - 非前缀推荐词
 - 异地场景
 - 附近Wifi
 - 排行榜、专题页
 - Tag挖掘和Tab梳理
 - 候选词归一化
 - 排序模型
 - 用户画像
 - 商家别名
 - 相关搜索多样性、半转义
 - 本地化文案
 - 决策信息组织
 - ...

流量占比的变化

QV占比



QV占比



■ 热词 ■ 搜索历史 ■ 智能补全 ■ 手工输入 ■ 相关搜索 ■ 其他

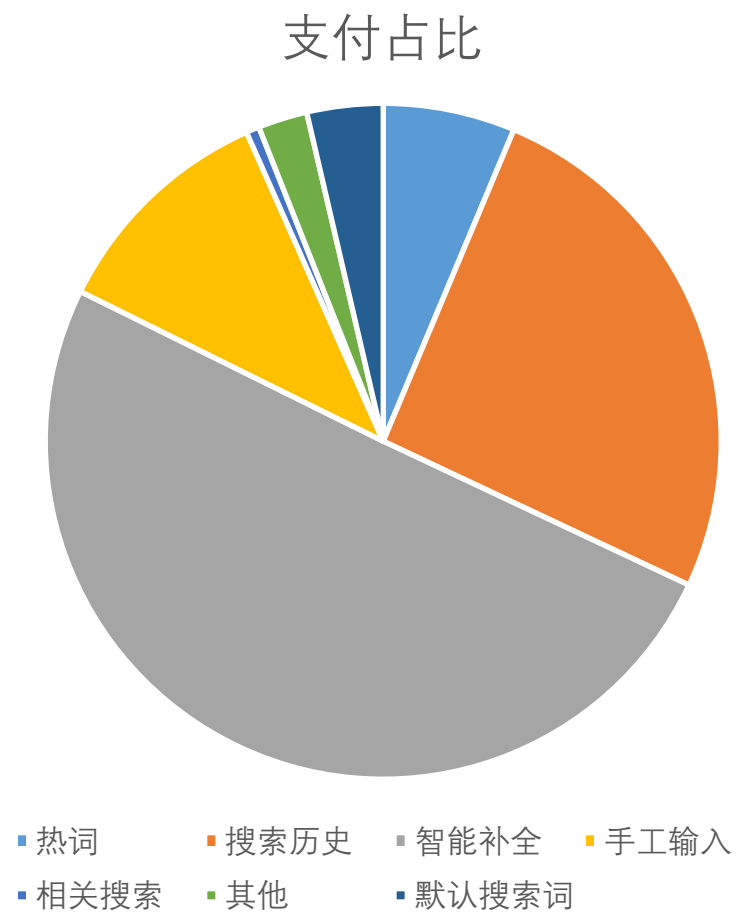
2016年1月

■ 热词 ■ 搜索历史 ■ 智能补全 ■ 手工输入 ■ 相关搜索 ■ 其他 ■ 默认搜索词

完整手工输入占比绝对值降低了15.13%

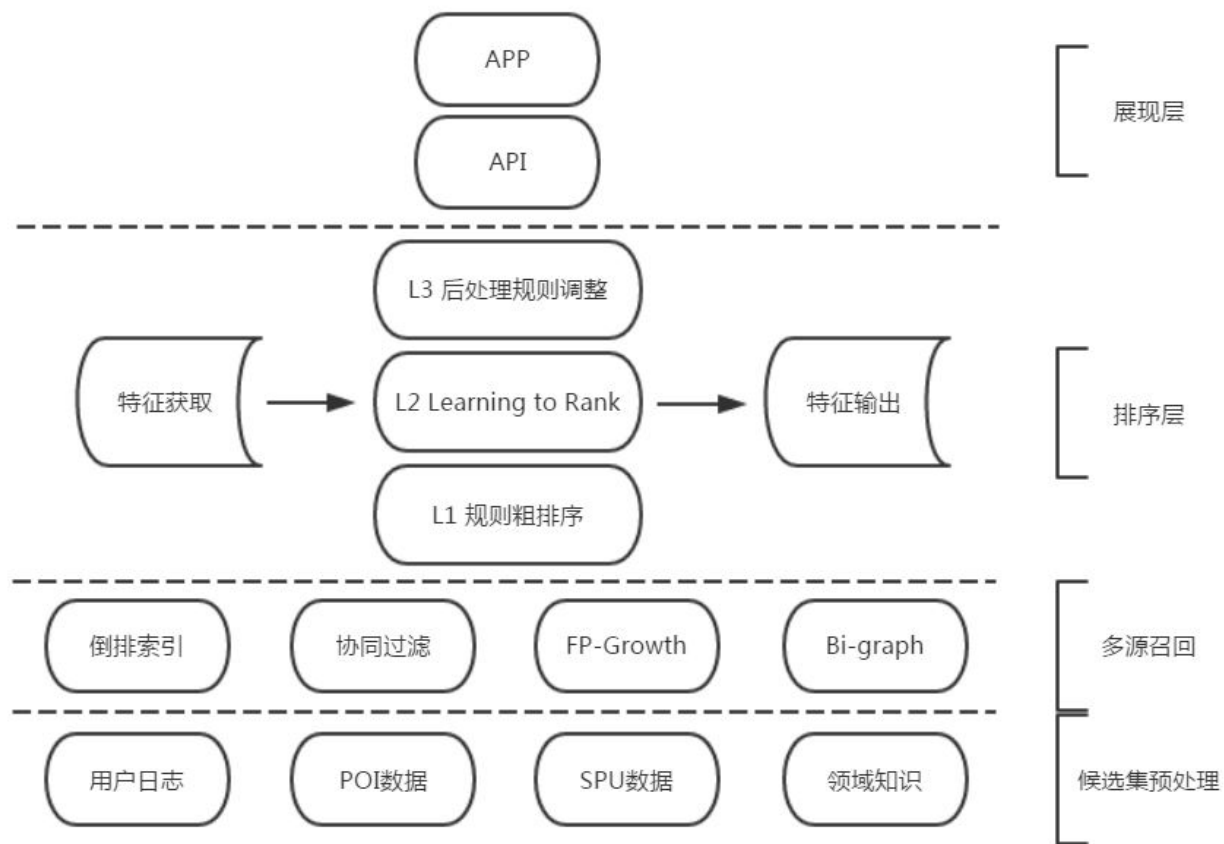
2016年12月

流量占比的变化



完整手工输入发起搜索的交易额仅占11.06%

架构升级-服务化



架构优化不会带来直接的效果提升，但是能加速迭代，解放人力。

将**多源召回层**和**排序层**做成通用服务，不同的产品通过**参数**控制模型的选择，从而将召回和排序的服务、流程、数据得以统一管理。

总结与思考

回顾过去，畅想未来

一些问题

指标漂移问题：从搜索入口到结果页，经历了漫长复杂的检索过程。引导的产品策略没问题，但是搜索结果不好，有这种可能。

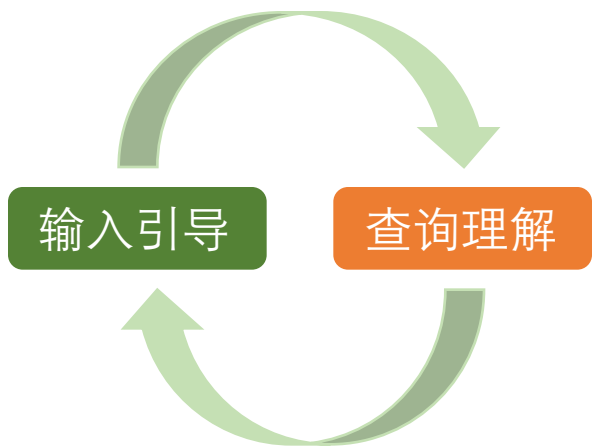
一种滞后的补救措施，对引导流量的低转化部分设置监控，定期评测，通过数据驱动查询理解优化。

出现**CTR上升**，**CVR下降**的情况怎么办？

看访购率，**流量**和**转化**两方面的衡量确定最终是否上线，但某一指标下降过大，都要慎重。转化更重要，马太效应一旦形成就很难逆转。

分流问题：**水平分层**是必须的，增加实验的并行化。**圈用户分流**是必须的，评测垂直叠加的优化。绝对准确难做到，至少减少相对误差。

一些思考



用户查询的输入引导和查询理解是互相促进的关系。

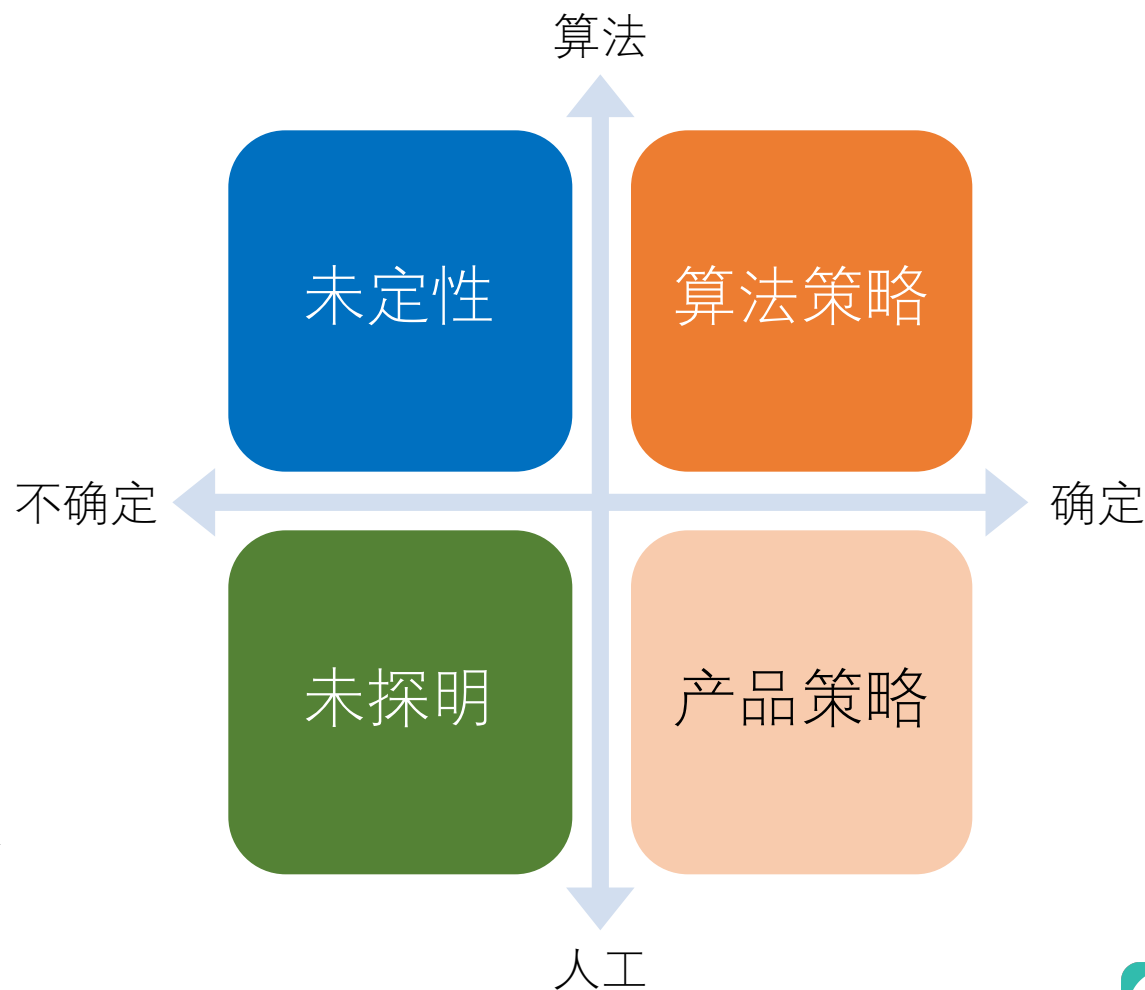
本质思想是将未知的搜索问题转化为已知问题。

用户引导产品在交互上互补、在场景上互补，形成一整套体系。

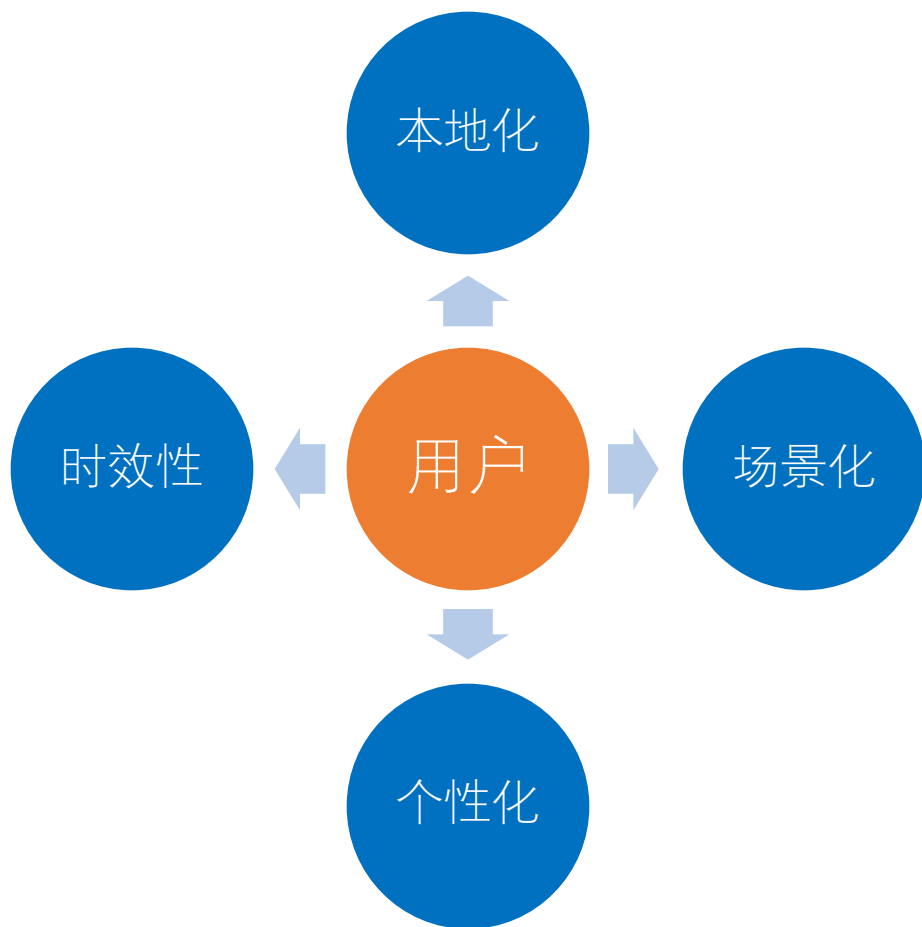
给用户选项，但别替用户做决策。

问题域的四象限

- 算法策略
 - 明确指标和方向，技术主导，深入优化
- 产品策略
 - 探寻方向，数据分析，产品实验
- 未定性
 - 技术主导，线上实验，做决策
- 未探明
 - 产品主导，小成本实验，确认进一步投入的可能



总结



驱动：
确定目标，通过数据与评测驱动。做好发现机制。

核心：
用户行为分析。换位思考，多想多看。

算法优化：
小步快跑，小成本的快速试验。

产品优化：
从数据分析明确想要证实的事情。

工程优化：
平台化、服务化、参数化。



谢谢！



欢迎加入美团搜索！



微信扫码查看详情

分享到朋友圈，告诉你的朋友这可能是最棒的技术沙龙